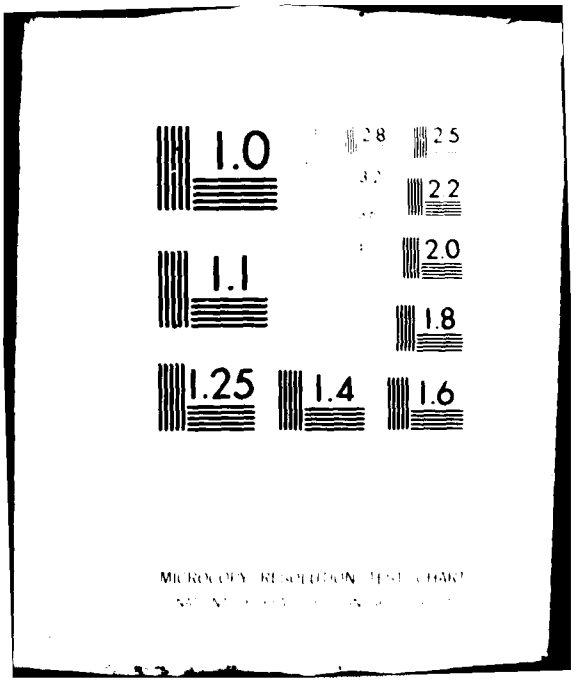
AD-A091 026 FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS F/G 12/1
THE INFLUENCE OF THE SAMPLE ON THE POSTERIOR DISTRIBUTION.(U)
APR 80 S FAHMY, C A PEREIRA, F PROSCHAN AFOSR-78-3678
UNCLASSIFIED FSU-STATISTICS-M545 AFOSR-TR-80-0995 NL

END
DATE FILMED
11-80
DTIC


1.0
1.1
1.25
1.4
1.6
2.0
1.8
2.2
2.5

AFOSR-TR-80-0995

LEVEL II

AD A091026

# THE INFLUENCE OF THE SAMPLE ON THE POSTERIOR DISTRIBUTION.

by

Salwa Fahmy[1], Carlos A. B. Pereira[2], and Frank Proschan[3]

FSU Statistics Report M545
AFOSR Technical Report No. 78-107

April 1980

The Florida State University
Department of Statistics
Tallahassee, Florida 32306

DTIC ELECTE OCT 30 1980 S D B

DDC FILE COPY

Key Words: Bayesian Inference, Monotone Likelihood Ratio Property, Multivariate Totally Positive Functions, Prior Distribution, Posterior Distribution.

[1,3]Research sponsored by the Air Force Office of Scientific Research, AFSC, USAF, under Grant AFOSR-78-3678.

[2]Research sponsored by CAPES and USP-Brasil.

Approved for public release; distribution unlimited.

80 10 6 104

# The Influence of the Sample on the Posterior Distribution

by

Salwa Fahmy, Carlos Pereira, and Frank Proschan


## ABSTRACT

In this paper, we present conditions on the likelihood function and on the prior distribution which permit us to assess the effect of the sample on the posterior distribution. Our work is inspired by Whitt (1979) J. Amer. Statist. Assoc. 74, and is based on the notion of multivariate totally positive functions introduced by Karlin and Rinott (1980), unpublished report.



| Accession For | |
|---|---|
| NTIS GRA&I | ☑ |
| DTIC TAB | ☐ |
| Unannounced | ☐ |
| Justification | |
| By | |
| Distribution/ | |
| Availability Codes | |
| Dist | Avail and/or Special |
| A | |

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is approved for public release IAW AFR 190-12 (7b).
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer

## 1. Introduction.

Whitt (1979) shows that in a Bayesian analysis, under certain general conditions, the larger the observations obtained, the larger stochastically will be the appropriate parameter of the posterior distribution. His interesting results are developed for the case of a univariate parameter. In the present paper, we show that his key ideas may be extended to the multivariate parameter case. Our basic tool is multivariate total positivity of order 2, introduced and treated by Karlin and Rinott (1980).

In Section 3, we present sufficient conditions on the likelihood function and on the prior distribution under which the posterior distribution of a multivariate parameter is stochastically increasing in the sample observations. In Section 4, we show how these results may be applied to a number of well known multivariate distributions, such as the normal, the multinomial, the negative multinomial, etc.

In a subsequent paper, we will extend Whitt's original notion to show that other properties of the data are "inherited" stochastically by the parameter posterior distribution.

## 2. Preliminaries.

In this section we present definitions, notation, and basic facts used throughout the paper.

2.1. Definition. A random vector $\underline{X}$ is said to be stochastically increasing in a random vector $\underline{Y}$ if $E[\phi(\underline{X})|\underline{Y} = \underline{y}]$ is increasing in $\underline{y}$ for every increasing function $\phi$ for which the expectation exists.

(A function $\phi$ on a subset of $R^k \to R$ is said to be increasing if it is increasing in each of its arguments).

Let $X_1, X_2, \ldots, X_k$ be totally ordered spaces. Let $X = X_1 \times X_2 \times \ldots \times X_k$. For every $\underline{x}, \underline{y} \in X$, define:

$$\underline{x} \vee \underline{y} = (\max(x_1, y_1), \ldots, \max(x_k, y_k))$$

$$\underline{x} \wedge \underline{y} = (\min(x_1, y_1), \ldots, \min(x_k, y_k)).$$

The following definitions are due to Karlin and Rinott (1980).

2.2. Definition. Consider a function $f(\underline{x})$ defined on X. We say that $f(\underline{x})$ is multivariate totally positive of order 2 if:

$$f(\underline{x} \vee \underline{y})f(\underline{x} \wedge \underline{y}) \geq f(\underline{x})f(\underline{y})$$

for every $\underline{x}, \underline{y} \in X$; we write f is $MTP_2$.

The following is a multivariate generalization of the familiar monotone likelihood ratio property.

2.3. Definition. Let $f_1$ and $f_2$ be two probability densities on $R^k$. If

$$f_2(\underline{x} \vee \underline{y})f_1(\underline{x} \wedge \underline{y}) \geq f_1(\underline{x})f_2(\underline{y})$$

for every $\underline{x}, \underline{y} \in R^k$, then we say that $f_2 >_{TP_2} f_1$.

2.4. Theorem. Let $f_1$ and $f_2$ be probability densities on $R^k$ such that:

$$f_2 >_{TP_2} f_1.$$

Let $\phi$ be an increasing function on $R^k$. Then:

$$\int \phi(\underline{x}) f_1(\underline{x}) \leq \int \phi(\underline{x}) f_2(\underline{x}).$$

For a proof, see Karlin and Rinott (1980).

## 3. Theoretical Results.

In the sequel, let $\underline{\theta}$ be a parameter taking values in the space $\Theta = \Theta_1 \times \ldots \times \Theta_k$, where each $\Theta_i$ is totally ordered. Let $\underline{X}$ be an observable random vector. Let $\Pi(\underline{x}|\underline{\theta})$ denote the conditional p.d.f. (or probability mass function) of $\underline{X}$ given $\underline{\theta}$ and let $\Pi(\underline{\theta})$ be the prior p.d.f. (or probability mass function) of $\underline{\theta}$. Finally, let $\Pi(\underline{\theta}|\underline{x})$ be the posterior p.d.f. (or probability mass function).

3.1. Theorem. Suppose

a) $\Pi(\underline{x}|\underline{\theta}) = h(\underline{x})g(\underline{x}, \underline{\theta})$, where $g(\underline{x}, \underline{\theta})$ is $MTP_2$ on $R^n \times \Theta$, and

b) $\Pi(\underline{\theta})$ is $MTP_2$ on $\Theta$. (3.1)

Then

$$\Pi(\underline{\theta}|\underline{x}^{(2)}) \geq_{TP_2} \Pi(\underline{\theta}|\underline{x}^{(1)})$$

for every $\underline{x}^{(1)}, \underline{x}^{(2)} \in R^n$ such that $\underline{x}^{(1)} \leq \underline{x}^{(2)}$.

Proof. The posterior density is given by:

$$\Pi(\underline{\theta}|\underline{x}) = C(\underline{x})g(\underline{x}, \underline{\theta})\Pi(\underline{\theta}),$$

where:

$$C(\underline{x}) = \left[\int_{\Theta} g(\underline{x}, \underline{\theta})\Pi(\underline{\theta})d\underline{\theta}\right]^{-1}$$

Let $\underline{x}^{(1)} \leq \underline{x}^{(2)}$. Define:

$$f_1(\underline{\theta}) = \Pi(\underline{\theta}|\underline{x}^{(1)}), \quad f_2(\underline{\theta}) = \Pi(\underline{\theta}|\underline{x}^{(2)}).$$

Let $\underline{\theta}, \underline{\nu} \in \Theta$. Write:

$$\begin{aligned} f_1(\underline{\theta})f_2(\underline{\nu}) &= C(\underline{x}^{(1)})g(\underline{x}^{(1)}, \underline{\theta})\Pi(\underline{\theta})C(\underline{x}^{(2)})g(\underline{x}^{(2)}, \underline{\nu})\Pi(\underline{\nu}) \\ &= C(\underline{x}^{(1)})C(\underline{x}^{(2)})g(\underline{x}^{(1)}, \underline{\theta})g(\underline{x}^{(2)}, \underline{\nu})\Pi(\underline{\theta})\Pi(\underline{\nu}) \\ &\leq C(\underline{x}^{(1)})C(\underline{x}^{(2)})g(\underline{x}^{(1)} \wedge \underline{x}^{(2)}, \underline{\theta} \wedge \underline{\nu})g(\underline{x}^{(1)} \vee \underline{x}^{(2)}, \\ &\qquad \underline{\theta} \vee \underline{\nu})\Pi(\underline{\theta} \wedge \underline{\nu})\Pi(\underline{\theta} \vee \nu), \end{aligned}$$

since both $g(\underline{x}, \underline{\theta})$ and $\pi(\underline{\theta})$ are $MTP_2$. Thus

$$f_1(\underline{\theta})f_2(\underline{\nu}) = C(\underline{x}^{(1)})g(\underline{x}^{(1)}, \underline{\theta} \wedge \underline{\nu})\pi(\underline{\theta} \wedge \underline{\nu})C(\underline{x}^{(2)})g(\underline{x}^{(2)}, \underline{\theta} \vee \underline{\nu})\pi(\underline{\theta} \vee \underline{\nu})$$

$$= f_1(\underline{\theta} \wedge \underline{\nu})f_2(\underline{\theta} \vee \underline{\nu}).$$

Hence

$$f_2 >_{TP_2} f_1. \ ||$$

3.2. Corollary. If (3.1) holds, then $\underline{\theta}$ is stochastically increasing in $\underline{X}$.

Proof. The result follows immediately from Theorems 2.4 and 3.1. ||

The conditions required to apply the previous Theorem 3.1 are too strong in many cases, and so we consider the following.

3.3. Theorem. Suppose:

$$\pi(\underline{x}|\underline{\theta}) = h(\underline{x}) \prod_{i=1}^{k} g_i(\theta_i, x_i)C(\underline{\theta}),$$

where $g_i(\theta_i, x_i)$ is $TP_2$ in $(\theta_i, x_i)$, $i = 1, \ldots, n$. (3.2)

Then:

$$\pi(\theta_j|X_1 = x_1, \ldots, X_j = x_j', \ldots, X_k = x_k) >_{TP_2} \pi(\theta_j|X_1 = x_1, \ldots, X_j = x_j, \ldots, X_k = x_k)$$

for every $x_j < x_j'$ and for every $j = 1, 2, \ldots, k$.

Proof. The marginal posterior density of $\theta_j$ is

$$\pi(\theta_j|\underline{x}) = h(\underline{x})g_j(\theta_j, x_j)\int\ldots\int \prod_{\substack{i=1 \\ i\neq j}}^{k} g_i(\theta_i, x_i)H(\underline{\theta}) \prod_{\substack{i=1 \\ i\neq j}}^{k} d\theta_i,$$

where $H(\underline{\theta}) = C(\underline{\theta})\pi(\underline{\theta})$. Define :

$$\psi(\theta_j, \underline{x}^{(j)}) = \int \cdots \int \prod_{\substack{i=1 \\ i \neq j}}^{k} g_i(\theta_i, x_i) H(\underline{\theta}) \prod_{\substack{i=1 \\ i \neq j}}^{k} d\theta_i,$$

where $\underline{x}^{(j)} = (x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k)$. Thus:

$$\Pi(\theta_j | \underline{x}) = h(\underline{x}) g_j(\theta_j, x_j) \psi(\theta_j, \underline{x}^{(j)}).$$

Let $x_j < x_j'$. Define:

$$f_1(\theta_j) = \Pi(\theta_j | X_1 = x_1, \ldots, X_j = x_j, \ldots, X_k = x_k),$$

$$f_2(\theta_j) = \Pi(\theta_j | X_1 = x_1, \ldots, X_j = x_j', \ldots, X_k = x_k).$$

Thus:

$$\frac{f_2(\theta_j)}{f_1(\theta_j)} = \frac{h(x_1, \ldots, x_j', \ldots, x_k) g_j(\theta_j, x_j')}{h(x_1, \ldots, x_j, \ldots, x_k) g_j(\theta_j, x_j)}$$

is increasing in $\theta_j$ by the $TP_2$ property of $g_j(\theta_j, x_j)$. Hence:

$$f_2 >_{TP_2} f_1. \quad ||$$

3.4. Remarks. Note that the result in Theorem 3.3 pertains to the marginal posterior density. Also, the result holds irrespective of which prior distribution we use. It follows from Theorems 2.4 and 3.3 that $E[\phi(\theta_j) | X_1 = x_1, \ldots, X_j = x_j, \ldots, X_k = x_k]$ is increasing in $x_j$ for every increasing function $\phi$.

If we consider the case where $\theta$ is of dimension two, we have the following result.

3.5. Theorem. Suppose

$$\Pi(\theta_1, \theta_2|x_1, x_2) = h(x_1, x_2) \prod_{i=1}^{2} g_i(\theta_i, x_i)C(\theta_1, \theta_2),$$

where $g_i(\theta_i, x_i)$ is $TP_2$ in $(\theta_i, x_i)$ for $i = 1, 2$, and $C(-\theta_1, \theta_2)$ is $TP_2$ in $(\theta_1, \theta_2)$.

Then for $x_2 < x_2'$,

$$\Pi(\theta_1|x_1, x_2) >_{TP_2} \Pi(\theta_1|x_1, x_2').$$

A similar result holds for the marginal posterior distribution of $\theta_2$.

Proof. The posterior density of $\theta_1$ is:

$$\Pi(\theta_1|x_1, x_2) = h(x_1, x_2)g_1(\theta_1, x_1)\int g_2(\theta_2, x_2)C(\theta_1, \theta_2)d\theta_2.$$

Define:

$$\psi(\theta_1, x_2) = \int g_2(\theta_2, x_2)C(\theta_1, \theta_2)d\theta_2.$$

Then $\psi(-\theta_1, x_2)$ is $TP_2$ in $(\theta_1, x_2)$ by the basic composition formula. [See Karlin, 1968, pp. 98-103.] Let $x_2 < x_2'$. Then:

$$\frac{\Pi(\theta_1|x_1, x_2)}{\Pi(\theta_1|x_1, x_2')} = \frac{h(x_1, x_2)\,\psi(\theta_1, x_2)}{h(x_1, x_2')\,\psi(\theta_1, x_2')},$$

which is increasing in $\theta_1$ by the $TP_2$ property of $\psi(-\theta_1, x_2)$.

3.6. Remark. Note that Theorem 3.5 involves conditions on the prior distribution. It follows from Theorems 2.4 and 3.5 that $E[\phi(\theta_1)|X_1 = x_1, X_2 = x_2]$ is decreasing in $x_2$ for every increasing function $\phi$.

## 4. Applications.

In this section, we present applications of the results of Section 3.

4.1. Multivariate Normal Distribution. Let $\underline{X}|\underline{\theta}$ be normally distributed with mean vector $\underline{\theta}$ and a known covariance matrix:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & 0 \\ & & \ddots & \\ 0 & & & \sigma_k^2 \end{pmatrix}.$$

Then, whatever prior we use, the conditions of Theorem 3.3 are satisfied. Hence from Remark 3.4, we conclude that $E[\phi(\theta_j)|X_1 = x_1, \ldots, X_j = x_j, \ldots, X_k = x_k]$ is increasing in $x_j$ for every increasing function $\phi$.

Consider the conjugate prior for $\underline{\theta}$, namely the multivariate normal distribution with mean vector $\underline{\mu}$ and covariance matrix D. Let $d^{ij}$ denote the $(i, j)^{th}$ element of $D^{-1}$. If $d^{ij} \leq 0$ for every $i \neq j$, then the conditions of Theorem 3.1 are satisfied. It follows from Cor. 3.2 that $\underline{\theta}$ is stochastically increasing in $\underline{X}$. If $k = 2$ and $d^{12} \geq 0$, then the conditions of Theorem 3.5 are satisfied. It follows that $E[\phi(\theta_1)|X_1 = x_1, X_2 = x_2]$ is decreasing in $x_2$.

4.2. Multinomial Distribution. Let $\underline{X}|\underline{\theta}$ have the multinomial distribution with parameters n and $\underline{\theta}$. Then

$$\Pi(\underline{x}|\underline{\theta}) = h(\underline{x}) \prod_{i=1}^{k} \theta_i^{x_i} I\left(\sum_{i=1}^{k} \theta_i = 1\right),$$

where

$$I\left(\sum_{i=1}^{k} \theta_i = 1\right) = \begin{cases} 1 & \text{if } \sum_{i=1}^{k} \theta_i = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Now $\theta_i^{x_i}$ is $TP_2$ in $(\theta_i, x_i)$. Hence the conditions of Theorem 3.3 are satisfied. It follows that for any prior, $E[\phi(\theta_j)|X_1 = x_1, \ldots, X_j = x_j, \ldots, X_k = x_k]$ is increasing in $x_j$ for every increasing function $\phi$.

If $k = 2$ and we use the conjugate prior for $\underline{\theta}$, namely the Dirichlet distribution with parametric vector $\underline{\alpha} = (\alpha_1, \alpha_2)$, the posterior density is

$$\Pi(\theta_1, \theta_2|x_1, x_2) = h(x_1, x_2) \prod_{i=1}^{2} \theta_i^{\alpha_i + x_i - 1} I(\theta_1 + \theta_2 = 1).$$

Since $I(-\theta_1 + \theta_2 = 1)$ is $TP_2$ in $(\theta_1, \theta_2)$, then by Theorem 3.5, $E[\phi(\theta_1)|X_1 = x_1, X_2 = x_2]$ is decreasing in $x_2$ for every increasing function $\phi$.

4.3. The Negative Multinomial. Let $\underline{X}|\underline{\theta}$ have the negative multinomial distribution with parameters $N$ and $\underline{\theta}$. Then

$$\Pi(\underline{x}|\underline{\theta}) = h(\underline{x}) \prod_{i=1}^{k} \theta_i^{x_i} I\left(\sum_{i=1}^{k} \theta_i \leq 1\right).$$

Hence we can conclude, using Theorem 3.3, that $E[\phi(\theta_j)|X_1 = x_1, \ldots, X_j = x_j, \ldots, X_k = x_k]$ is increasing in $x_j$ for every increasing function $\phi$.

If $k = 2$ and we use the conjugate prior for $\underline{\theta}$, namely the Dirichlet distribution, we can deduce that $E[\phi(\theta_1)|X_1 = x_1, X_2 = x_2]$ is decreasing in $x_2$ for every increasing function $\phi$.

4.4. The Multivariate Hypergeometric Distribution. Suppose

$$\Pi(\underline{x}|\underline{\theta}) = h(\underline{x}) \prod_{j=1}^{k} \binom{\theta_j}{x_j} I\left(\sum_{i=1}^{k} \theta_i = N\right).$$

Since $\binom{\theta_j}{x_j}$ is $TP_2$ in $(\theta_j, x_j)$, it follows by Theorem 3.3 that for any increasing function $\phi$, $E[\phi(\theta_j)|X_1 = x_1, \ldots, X_j = x_j, \ldots, X_k = x_k]$ is increasing in $x_j$. If $k = 2$, Theorem 3.5 applies and it is easy to show that Theorem 2 in Whitt (1979) follows immediately.

4.5. Uniform Distribution. Suppose $T_1, T_2, \ldots, T_n$ is a random sample from a uniform distribution on the interval $(\theta_1, \theta_2)$. Let $X_1 = \min(T_1, T_2, \ldots, T_n)$, $X_2 = \max(T_1, T_2, \ldots, T_n)$. Suppose also that we use the conjugate prior, namely the bilateral bivariate Pareto distribution [see DeGroot, (1970), pp. 172-174]. It is easy to show that Theorems 3.3 and 3.5 apply and so we conclude that $E[\phi(\theta_1)|X_1 = x_1, X_2 = x_2]$ is increasing in $x_1$ and decreasing in $x_2$ for every increasing function $\phi$.

4.6. The Exponential Family of Distributions. We obtain an application of a general nature of the results of this paper. Let $\underline{X}|\underline{\theta}$ belong to the exponential family, reparametrized, so that:

$$\Pi(\underline{x}|\underline{\theta}) = h(\underline{x})e^{\underline{\theta}'\underline{x}}C(\underline{\theta}).$$

Since $e^{\theta_i x_i}$ is $TP_2$ in $(\theta_i, x_i)$, it follows by Theorem 3.3 that for any increasing function $\phi$, $E[\phi(\theta_j)|X_1 = x_1, \ldots, X_j = x_j, \ldots, X_k = x_k]$ is increasing in $x_j$.

## REFERENCES


1 DeGroot, Morris H. (1970). Optimal Statistical Decisions. McGraw-Hill Book Co., New York.

2 Karlin, S. (1968). Total Positivity. Stanford University Press.

3 Karlin, S. and Rinott, Y. (1980). Classes of Orderings of Measures and Related Correlation Inequalities. I. Multivariate Totally Positive Distributions. Unpublished report.

4 Karlin, S. and Rinott, Y. (1980). Classes of Orderings of Measures and Related Correlation Inequalities. II. Multivariate Reverse Rule Distributions. Unpublished report.

5 Whitt, W. (1979). A Note on the Influence of the Sample on the Posterior Distribution. J. Amer. Statist. Assoc. 74 424-426.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

| 1. REPORT NUMBER | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
|---|---|---|
| AFOSR-TR- 80-0995 | AD-A090 026 | |

4. TITLE

The Influence of the Sample on the Posterior Distribution

5. TYPE OF REPORT & PERIOD COVERED
INTERIM

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)

Salwa Fahmy
Carlos A. B. Pereira
Frank Proschan

8. CONTRACT OR GRANT NUMBER(s)
AFOSR 78-3678

9. PERFORMING ORGANIZATION NAME & ADDRESS

The Florida State University
Department of Statistics
Tallahassee, Florida 32306

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
61102F
2304/A5

11. CONTROLLING OFFICE NAME & ADDRESS

Air Force Office of Scientific Research/NM
Bolling Air Force Base, D.C. 20332

12. REPORT DATE
April, 1980

13. NUMBER OF PAGES
11

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS

Bayesian Inference, Monotone Likelihood Ratio Property, Multivariate Totally Positive Functions, Prior Distribution, Posterior Distribution

20. ABSTRACT

In this paper, we present conditions on the likelihood function and on the prior distribution which permit us to assess the effect of the sample on the posterior distribution. Our work is inspired by Whitt (1979) J. Amer. Statist. Assoc. 74, and is based on the notion of multivariate totally positive functions introduced by Karlin and Rinott (1980), unpublished report.

UNCLASSIFIED